#6

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/11, 15/81, C07K 14/395, C12N 1/19 — A2

(11) International Publication Number: **WO 96/32475**

(43) International Publication Date: 17 October 1996 (17.10.96)

(21) International Application Number: PCT/US96/04783

(22) International Filing Date: 10 April 1996 (10.04.96)

(30) Priority Data:
08/422,107 12 April 1995 (12.04.95) US

(71) Applicant: UNIVERSITY OF WASHINGTON [US/US]; Suite 200, 1107 N.E. 45th Street, Seattle, WA 98105 (US).

(71)(72) Applicant and Inventor: CHENG, Cheng [CN/CA]; 4246 Union, Burnaby, British Columbia V5C 2X4 (CA).

(72) Inventor: YOUNG, Elton, T.; 2617 Boyer Avenue East, Seattle, WA 98102 (US).

(74) Agent: PARKER, Gary, E.; ZymoGenetics, Inc., 1201 Eastlake Avenue East, Seattle, WA 98102 (US).

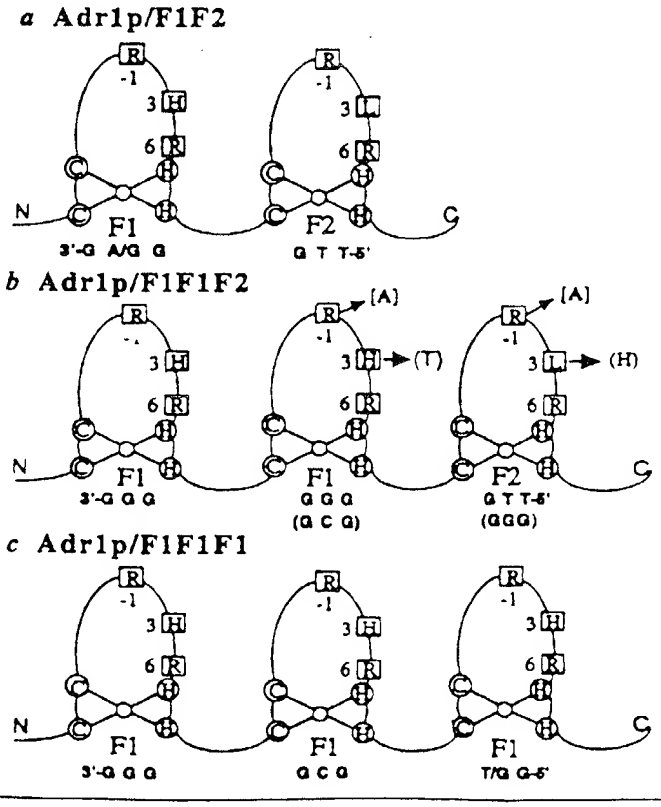
(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: METHODS FOR PREPARING DNA-BINDING PROTEINS

(57) Abstract

Methods for preparing DNA-binding proteins having altered binding specificity are disclosed. The binding specificity of a parent DNA-binding protein comprising first and second Cys_2-His_2 zinc fingers is altered by the addition of an additional zinc finger, wherein the altered specificity is a result of interactions between nucleotides in a target sequence and amino acid residues in each of the first, second and additional zinc fingers. The altered DNA-binding proteins are useful within methods for preparing polypeptides.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

Description

METHODS FOR PREPARING DNA-BINDING PROTEINS

Government Support

This invention was made with government support under grant number GM26079 awarded by the National Institutes of Health. The government has certain rights in the invention.

Background of the Invention

A major area of interest in the field of molecular biology has been the identification of mechanisms by which gene expression is regulated. Through research in this area, it has been established that the primary control of gene expression lies at the level of gene transcription. Cells respond to intracellular and extracellular cues by turning certain genes on or off, and by modulating the level of transcription of active genes. Many genes in multicellular organisms are transcribed only in particular tissues where their protein products are required (Darwell, Nature 297:365-371, 1982; Latchman, Gene Regulation: A Eucaryotic Perspective, Unwin Hyman, London, 1990). Genes that are regulated in parallel in response to a particular inducing signal or in a particular tissue appear to contain common DNA sequence elements which are often, but not always, located upstream of the start site of transcription (Maniatis et al., Science 236:1237-1245, 1987). These elements, which provide recognition sites for protein transcription factors, play an essential role in the expression of genes.

It is believed that transcription factors have increased in number and diversity during evolution by processes such as gene duplication, divergence and exon

shuffling (Mitchell et al., Science 244:371-378, 1989). With the increasing complexity of nuclear DNA sequences that accompanied and presumably accounted for evolutionary changes, internal regions in small DNA-binding motifs may have been duplicated to create DNA-binding proteins with greater specificity.

One well-characterized family of DNA-binding proteins is the Cys_2-His_2 class of "zinc finger" proteins, which constitute one of the major classes of DNA-binding proteins in eukaryotes (Berg, Proc Natl. Acad. Sci. USA 85:91-102, 1988). The zinc finger motif was so named because of the tandemly repeating pattern observed in the amino acid sequence of the RNA polymerase III transcription factor TFIIIA (Miller et al., EMBO J. 4:1609, 1985). This motif includes approximately twenty-eight to thirty amino acid residues, with two cysteine and two histidine residues serving to stabilize the domain structure by tetrahedral coordination of a single Zn^{2+} ion. A region of approximately twelve amino acids between the invariant cysteine-histidine pairs is characterized by scattered basic residues and several conserved hydrophobic residues. Zinc finger motifs can be represented by the consensus sequence Pro-(Tyr/Phe)-Xaa-Cys-Xaa_{2-4}-Cys-Xaa-Xaa-Xaa-Phe-Xaa-Xaa^1-Xaa-Xaa-Xaa^2-Leu-Xaa-Xaa^3-His-Xaa_{3-4}-His-Thr-Gly-Glu-Lys (SEQ ID NOS:1-6), wherein each Xaa is individually a variable amino acid (i.e. each Xaa may be different), subscripts indicate the number of variable amino acids, and superscript numbers denote amino acid residues forming the predominant sequence-specific contacts with DNA. Zinc finger proteins comprise tandem arrays of these motifs. Individual fingers within the array generally associate with three-base-pair subsites in DNA, with the target DNA triplets being contiguous but not usually overlapping. Studies of crystal structure indicate that only one of the two DNA strands is the major contributor to specific binding. The predominant sequence-specific contacts (designated Xaa^1, Xaa^2 and Xaa^3

above) are located at positions -1, 3 and 6 relative to an alpha helical region of the polypeptide backbone. The orientation of protein and DNA is antiparallel, that is the N-terminal to C-terminal orientation of the contact amino acid residues is antiparallel to the 5'→3' orientation of the contacted DNA strand:

```
amino acid:    (N)Arg-//-His-//-Arg(C)
                   |      |      |
DNA:           (3')G      A      G(5').
```

Multiple examples of this zinc finger motif have been identified in a number of transcription factors, including Sp1 (Kadonaga et al., Cell 51:1079-1090, 1987), the Kruppel protein (ReDemann et al., Nature 332:90-92, 1988), the yeast ADR1 transcription factor (Hartshorne et al., Nature 320: 283-287, 1987) and Zif268 (Christy et al., Proc. Natl. Acad. Sci. USA 85:7857-7861, 1988).

The tertiary structures of zinc finger motifs are essentially the same within and among proteins. Conserved structural features include the relative position and size of the alpha helix and the position of the turn between the alpha helix and the reverse beta sheet.

Zinc finger proteins form a subset of the transcription factors, proteins that interact with genes to modulate transcription of those genes. Transcriptional regulation results from the combined effects of a number of factors acting in concert to determine the frequency of transcription initiation. Zinc finger-type transcription factors can be classified as transcriptional activators or repressors. Transcriptional activators are believed to function by interacting, directly or indirectly, with components of the basal transcription complex to enhance formation of a pre-initiation complex. Transcriptional repressors are believed to function by altering chromatin structure, by preventing assembly of basal transcription factors, or by inhibiting the function of transcriptional activators. Activators and repressors generally contain

both DNA-binding and regulatory domains. Families of activators and repressors can be defined by shared structural characteristics (e.g., an acidic or glutamine-rich activation domain, an alanine-rich domain, or a Kruppel-associated box), although others do not exhibit obvious similarities to these groupings.

The manner in which DNA-binding proteins are able to recognize and bind to their specific target sequences is central to the control of differentiation and development. A knowledge of these processes would provide the opportunity to manipulate gene expression in such applications as recombinant protein production and gene therapy.

Intitial attempts to manipulate zinc finger motifs were motivated by a belief that a set of rules could be elucidated to predict binding specificity of any individual zinc finger or combination of fingers (see, e.g., Desjarlais and Berg, Proc. Natl. Acad. Sci. USA 89:7345-7349, 1992), thereby providing for the design of proteins capable of binding to any given sequence. To date, however, no such rules exist, and more recent studies, such as those with WT1 disclosed by Drummond et al. (Mol. Cell. Biol. 14:3800-3809, 1994) indicate that no generalizable code for DNA recognition by zinc finger proteins exists. Other studies of mutagenized zinc fingers (e.g., Nardelli et al., Nuc. Acids Res. 20:4137-4144, 1992; Thukral et al., Mol. Cell. Biol. 12:2784-2792, 1992) have shown that the context of a contact amino acid affects its ability to interact with DNA.

There remains a need in the art for methods of preparing DNA-binding proteins, including transcription factors, having desired specificities. There also remains a need in the art for methods of regulating gene transcription in genetically engineered cells. The present invention provides such methods as well as other, related advantages.

Summary of the Invention

The present invention is directed to modified DNA-binding proteins of the zinc finger type, and to methods of preparing and using such modified proteins and genes encoding them.

Within one aspect, the invention provides a method for preparing a DNA-binding protein having altered binding specificity. The method comprises the steps of (a) selecting a parent DNA-binding protein comprising first and second Cys_2-His_2 zinc fingers, wherein the DNA binding specificity of the parent protein is known; (b) adding an additional Cys_2-His_2 zinc finger to the parent protein to produce an altered DNA-binding protein; and (c) determining the DNA binding specificity of the altered DNA-binding protein, wherein the binding specificity of the altered protein is a result of interactions between nucleotides in a target sequence and amino acid residues in each of the first, second, and additional zinc fingers. Within one embodiment, the additional zinc finger is a duplicate of one of the first and second zinc fingers. Within another embodiment, the parent DNA-binding protein is a wild-type *Saccharomyces cerevisiae* ADR1 protein or MIG1 protein. Within an additional embodiment, the parent DNA-binding protein is *Saccharomyces cerevisiae* ADR1 having a mutation in one of the first or second fingers that changes DNA binding specificity as compared to wild-type *Saccharomyces cerevisiae* ADR1. Within yet another embodiment, the target sequence is from 9 to 15 nucleotides in length. Within a further embodiment, the altered DNA-binding protein has three, four or five Cys_2-His_2 zinc fingers, each of which interacts with the target sequence. Within another embodiment, the parent DNA-binding protein has two, three or four Cys_2-His_2 zinc fingers.

Within the methods of the present invention the step of determining the DNA binding specificity of the altered protein can be carried out *in vitro* or *in vivo*.

Within one embodiment, the determining step comprises measuring electrophoretic mobility of a complex of the altered DNA-binding protein and a DNA molecule. In another embodiment, the determining step comprises preparing a mixture of the altered DNA-binding protein and a DNA molecule comprising a predicted binding site under conditions suitable for formation of protein-DNA complexes, and measuring complex formation between the altered DNA-binding protein and the DNA molecule, such as by measuring electrophoretic mobility of a complex of the altered DNA-binding protein and the polynucleotide molecule. In another embodiment, the determining step comprises (a) preparing a mixture comprising the altered DNA-binding protein and a plurality of DNA molecules under conditions suitable for complex formation between the altered protein and a target DNA molecule, (b) isolating a complex of the altered DNA-binding protein and a target DNA molecule, (c) amplifying the target DNA molecule, and (d) determining the sequence of a binding site for the altered DNA-binding protein in the target DNA molecule. Within an alternative embodiment the determining step comprises (a) culturing a first cell into which has been introduced a first DNA construct comprising a reporter gene operably linked to a transcription promoter segment containing a potential target sequence, (b) culturing a second cell into which has been introduced the first DNA construct and a second DNA construct that directs expression of the altered DNA-binding protein, and (c) measuring transcription of the reporter gene in the first and second cells, wherein a difference in relative transcription levels is indicative of binding of the altered DNA-binding protein to the potential target sequence.

Within another aspect of the invention there is provided a DNA-binding protein comprising first and second Cys_2-His_2 zinc fingers which has been modified to contain an additional Cys_2-His_2 zinc finger, wherein the DNA-

binding protein binds to a binding site in DNA, wherein the binding is a result of interactions between nucleotides in the binding site and amino acid residues in each of the first, second, and additional zinc fingers. Within one embodiment, the DNA-binding protein contains only three zinc fingers. Within another embodiment, the additional zinc finger is a duplicate of one of the first and second zinc fingers. Within another embodiment, the DNA-binding protein is a *Saccharomyces cerevisiae* protein selected from the group consisting of ADR1 and MIG1. Within an additional embodiment, the DNA-binding protein is a *Saccharomyces cerevisiae* ADR1 protein which has been modified to contain a third Cys_2-His_2 zinc finger, wherein the ADR1 protein binds to a binding site other than TTGG(A/G)G as a result of interactions between nucleotides in the binding site and amino acid residues in the third zinc finger. The third zinc finger can be a duplicate of a zinc finger in a wild-type ADR1 protein. In an alternative embodiment, the third zinc finger is that of a DNA-binding protein other than ADR1. In a further embodiment, the ADR1 protein is further modified to contain a fourth Cys_2-His_2 zinc finger, wherein the fourth zinc finger alters binding specificity of the ADR1 protein.

Within a third aspect of the invention there is provided a cultured eukaryotic cell into which has been introduced a gene encoding a DNA-binding protein comprising first and second Cys_2-His_2 zinc fingers, wherein the gene has been modified so that the DNA-binding protein contains an additional Cys_2-His_2 zinc finger, wherein the DNA-binding protein binds to a binding site in DNA as a result of interactions between nucleotides in the binding site and amino acid residues in each of the first, second, and additional zinc fingers. Within one embodiment, the cultured eukaryotic cell is a fungal cell, such as a yeast cell or a filamentous fungal cell. Within another embodiment, the cultured eukaryotic cell is a

Saccharomyces cerevisiae cell. Within an additional embodiment, the DNA-binding protein is a modified *S. cerevisiae* ADR1 or MIG1 protein. Within another embodiment, a first DNA segment encoding a polypeptide of interest operably linked to a second DNA segment comprising a transcription promoter and a binding site for the DNA-binding protein are introduced into the cell, wherein binding of the DNA-binding protein to the binding site stimulates transcription of the first DNA segment.

A further aspect of the invention provides a method for preparing a polypeptide of interest comprising the steps of (a) culturing a yeast cell into which has been introduced (i) an *ADR1* gene modified to encode a protein containing a third Cys_2-His_2 zinc finger, wherein the *ADR1*-encoded protein binds to a binding site other than TTGG(A/G)G as a result of interactions between nucleotides in the binding site and amino acid residues in the third zinc finger; and (ii) a first DNA segment encoding the polypeptide of interest operably linked to a second DNA segment comprising a transcription promoter and a binding site for the *ADR1*-encoded protein, wherein binding of the *ADR1*-encoded protein to the binding site stimulates transcription of the first DNA segment, under conditions suitable for expression of the *ADR1* gene and the first DNA segment; and (b) isolating the polypeptide of interest from the yeast cell.

Another aspect of the invention provides a cultured eukaryotic cell into which has been introduced a gene encoding a chimeric transcription factor comprising a *S. cerevisiae* ADR1 DNA-binding domain modified to contain a third Cys_2-His_2 zinc finger, wherein the ADR1 DNA-binding domain binds to a binding site other than TTGG(A/G)G as a result of interactions between nucleotides in the binding site and amino acid residues in the third zinc finger, and wherein the chimeric transcription factor further comprises a non-ADR1 transcription activation or repression domain operably linked to the DNA-binding

domain. Within one embodiment, the non-ADR1 domain is a transcription activation domain from *S. cerevisiae* GAL4, *S. cerevisiae* GCN4, or human or mouse SP1. Within another embodiment, the non-ADR1 domain is a steroid receptor family transcription activation domain.

These and other aspects of the invention will become evident upon reference to the following detailed description and the attached drawing.

Brief Description of the Drawing

The Figure illustrates the structures of wild-type and altered ADR1 zinc-finger domains and their binding sites. Standard single-letter abbreviations are used for amino acid residues and nucleotides.

Detailed Description of the Invention

Before describing the invention in detail, it may be helpful to an understanding thereof to define certain terms used herein. The term "gene" is used herein to describe a DNA segment that encodes a polypeptide or protein. The term encompasses naturally occuring and synthetic DNAs (including cDNAs), as well as copies of such molecules. Genes may or may not include non-coding sequences such as introns, promoters, and other flanking sequences. The term "modified", when used herein to describe genes and proteins, indicates that the material has been changed by human intervention so that it differs from the respective parent material. Such modified genes and proteins include the originally modified material as well as copies thereof. A protein may be modified by expressing a modified gene in a cultured cell.

The present invention provides methods for altering DNA-binding specificities of proteins containing Cys_2-His_2 zinc fingers. Within these methods, one or more additional zinc fingers are added to a parent zinc finger protein. The present inventors have found that additional zinc fingers can combine with existing zinc fingers in a

protein to provide a new binding specificity wherein amino acid residues in each of the fingers interact with nucleotides in a target DNA sequence so as to alter the binding specificity of the parent protein. Within one embodiment of the invention, altered *in vivo* binding specificity of the zinc finger proteins is exploited to regulate expression levels of cloned DNAs.

As noted above, the zinc finger motif is characterized by an approximately 28-residue consensus sequence shown in SEQ ID NOS:1-6. Those skilled in the art will recognize that not all zinc finger motifs precisely conform to this consensus sequence. It is conventional in the art to portray the DNA-binding residues of a zinc finger as a three-residue sequence (e.g. Arg-His-Arg), even though these residues are not contiguous within the protein.

The methods of the present invention begin with a selected parent DNA-binding protein comprising first and second Cys_2-His_2 zinc fingers. The DNA binding specificity of the parent protein is known, either through prior experimentation or by the application of experimental techniques known in the art, such as DNA migration assays (Carey, J. in Sauer, R.T. (ed.), Methods Enzymol. 208:103-117, 1991), binding site selection and amplification (Blackwell and Weintraub, Science 250: 1104-1110, 1990), and phage display ("panning") (Rebar and Pabo, Science 263: 671-673, 1994; Jamieson et al., Biochemistry 33: 5689-5695, 1994). Within a typical procedure, a protein or polypeptide comprising a zinc finger DNA-binding domain is incubated with a plurality of radiolabeled DNA probes under conditions that promote binding of zinc fingers to their recognition sequences within DNA. Typical incubation conditions are 25 mM HEPES pH 7.5, 5 mM $MgCl_2$, 10 μM $ZnCl_2$, 1 mM dithiothreitol, 50 mM KCl, 1 μg/μl dI-dC, 10% glycerol. The binding reaction is allowed to proceed for approximately ten minutes, after which it is electrophoresed on a polyacrylamide gel. A

change in the mobility of a labeled probe is indicative of binding. Table 1 lists examples of known zinc finger proteins and their binding specificities.

Table 1

Protein	Binding Site (5'→3')	Ref.
ADR1	TTGG(A/G)G	1
Krox20	GCGGGGGCG	2
Sp1	(G/T)GGGCGG(G/A)(G/A)	3
BrlA	CAAGGGG	4
TFIIIA	GGnnGGnAGGAnnGGGnGGnnnAnnnG (SEQ ID NO:7)	5
Evi-1	GA(C/T)AAGATAAGATAA (SEQ ID NO:8)	6
MIG1	(G/C)(C/T)GG(G/A)G	7
TTK	TAAGGAA	8

References:

1. Thukral et al., Mol. Cell. Biol. 12:2784, 1992.
2. Chavrier et al., EMBO J. 9: 1209-1218, 1990.
3. Kadonaga et al., Cell 51:1079, 1987.
4. Chang et al., Genetics 133:29, 1992.
5. Fairall et al., J. Mol. Biol. 192:577, 1986.
6. Delwel et al., Mol. Cell. Biol. 13:4291, 1993.
7. Lundin et al., Mol. Cell. Biol. 14:1979, 1994; Nehlin et al., EMBO J. 9:2891, 1990.
8. Fairall et al., Nature 366:483, 1993.

Those skilled in the art will recognize that Table 1 lists only a subset of known zinc finger proteins, and that mutants and engineered variants of many zinc finger proteins and zinc finger domains have been characterized. See, for example, Thiesen and Bach, FEBS Lett. 283:23-26, 1991; Desjarlais and Berg, Proc. Natl. Acad. Sci. USA 89:7345-7349, 1992; Nardelli et al., Nuc. Acids Res. 20:4137-4144, 1992; Warriar et al., J. Biol. Chem. 269:29016-29023, 1994; and Wang et al., J. Biol. Chem. 269:10771-10775, 1994.

Although binding sites for zinc finger proteins are commonly shown as nucleotide triplets, not all bases within a triplet must contribute equally to binding, and individual zinc fingers within a protein may bind one, two or three bases on the contact strand of the target DNA. Experimental evidence suggests that a two-base binding site is most common. Thus, some bases within binding sites can be varied without complete loss of binding (and binding protein function).

The parent DNA-binding protein may be a native (i.e. naturally occuring) protein or a modified protein. Modified proteins are commonly produced through the application of genetic engineering techniques such as site-directed mutagenesis (e.g. Bahl, U.S. Patent No. 4,351,901; Zoller and Smith, DNA 3:479, 1984; Kunkel, Proc. Natl. Acad. Sci. USA 82:488-492, 1985), polymerase chain reaction (Mullis et al., U.S. Patent No. 4,683,195; Mullis, U.S. Patent No. 4,603,202) or the like. Modified proteins include those altered within one or more of the zinc fingers and those having sequence alterations elsewhere. In the former case, the alteration may or may not affect the DNA-binding specificity of the protein. These modified proteins include chimeric DNA-binding proteins wherein a native or altered DNA-binding domain is operably linked to a second domain from another protein, which second domain is an activator or repressor of transcription. For example, a native or altered DNA-binding domain of the *S. cerevisiae* ADR1 protein can be joined to the acidic activation domain of the herpes simplex virus protein VP16 (Triezenberg et al., Genes Dev. 2:718-729, 1988), the activation domain of *S. cerevisiae* GAL4 (Leuther et al., Cell 72:575-585, 1993) or GCN4 (Hope et al., Nature 333:635-640, 1988), an activation domain of a member of the steroid receptor family, or the glutamine-rich activation domain of human or mouse SP1 (Kadonaga et al., ibid.). The activation domain may be positioned N-terminal or C-terminal to the DNA-binding domain.

In addition to the first and second Cys_2-His_2 zinc fingers, the parent protein may have third, fourth and further zinc fingers. Such third, fourth and further zinc fingers may be naturally occuring in the protein or may be added to the protein as disclosed herein. The present invention thus provides, within one embodiment, methods for adding a plurality of zinc fingers to a DNA-binding protein.

After a parent DNA-binding protein is selected, an additional Cys_2-His_2 zinc finger is added to the parent protein. The additional zinc finge is added adjacent to one of the zinc fingers of the parent protein, that is the additional finger is connected to an existing finger by the conserved linker peptide found in zinc finger proteins. This linker is characterized by a five-residue consensus sequence, Thr-Gly/Asn-Glu-Lys-Pro (SEQ ID NO:9). The additional finger can thus be positioned between fingers of the parent protein or at either end of a plurality of contiguous fingers.

The methods of the present invention are of particular utility in producing altered DNA-binding proteins for use within genetically engineered protein production systems. It is therefore preferred to add the additional zinc finger by manipulating a DNA molecule encoding the parent protein, so that a DNA segment encoding the additional zinc finger is inserted into or joined to the DNA molecule encoding the parent protein. Suitable methods of manipulation in this regard include enzymatic digestion and ligation of DNA segments, loop-out mutagenesis using the polymerase chain reaction (PCR; see Mullis et al., U.S. Patent No. 4,683,195 and Mullis, U.S. Patent No. 4,683,202), and *de novo* synthesis of DNA molecules, as well as combinations of these methods. See, in general, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory, 1989. As discussed in more detail below, a DNA segment encoding a DNA-binding fragment of a DNA-binding protein

may be modified and subsequently ligated to the remainder of the protein coding sequence.

The additional zinc finger can be a duplicate of a zinc finger in the parent protein. In an alternative embodiment, the additional zinc finger has the binding specificity of a zinc finger in the parent protein but will differ in amino acid sequence. In a third embodiment, the additional zinc finger differs from the zinc fingers in the parent protein in both sequence and binding specificity. The present invention thus provides zinc finger proteins having multiple copies of zinc fingers or a plurality of different zinc fingers.

The binding specificity of the altered DNA binding protein is then determined. Binding specificity is defined as the ratio of K_{app} for a particular DNA sequence to K_{app} for a random DNA sequence. A protein will be considered to specifically bind a particular sequence when the ratio is at least 10. It is preferred that the ratio be at least 100. K_{app} (the apparent equilibrium constant) of DNA binding is calculated from the equation $[CD]/[D] = -K_{app}[CD] + K_{app}[C_0]$ (Baker et al., J. Biol. Chem. 261:5275-5282, 1986) and is determined by titration of zinc finger proteins using known amounts of DNA probes according to conventional techniques. Binding affinity is commonly determined by measuring the electrophoretic mobility of a complex of the protein and a double-stranded DNA. Briefly, DNA migration assays are carried out using a known amount of zinc finger protein or polypeptide and varying amounts of labeled target DNA. The protein or polypeptide is incubated with labeled DNA probes, then the incubated mixture is electrophoresed on a gel to allow detection of changes in the mobility of the DNA probes. Changes in mobility indicate the formation of a complex between the protein or fragment and DNA, allowing the calculation of affinities for different DNA sequences and ratios of affinities. The amounts of free DNA and DNA-protein complex are determined by measuring

the amounts of radioactivity in each of these components. Such measurements are made by methods known in the art, such as autoradiography followed by quantitative densitometry or phosphor image analysis of the gel. Predictions of DNA-binding specificity may be made on the basis of known specificities. Such predictions can be used to guide the selection of probes to be included in the incubated mixture.

Within another embodiment of the invention, a mixture of the altered DNA-binding protein or a DNA-binding fragment thereof and a plurality of double-stranded DNA molecules is prepared. The mixture is incubated under conditions suitable for complex formation between the altered protein and a target polynucleotide. Typically, the altered protein will be combined with a mixture of DNA molecules, such as a random sequence pool, and binding is detected using a gel mobility assay. When working with random pools, the number of bases in the DNA to be randomized will usually be equal to three times the number of variant zinc fingers in the protein, although if the binding properties of a variant finger or position within the finger are known, fewer positions can be randomized. For example, if only one variant finger is included, only three positions need be randomized. A complex of the altered DNA-binding protein and a target DNA is then isolated. The target DNA is separated from the protein (e.g., by extraction with phenol) and amplified by PCR. The consensus of several selected and amplified DNA sequences is the binding sequence of the DNA-binding protein. In general, this process of binding, isolation and amplification will be repeated using the PCR products of one round as the pool for the next round of selection and continuing until most of the sequences within the pool can be bound by the protein. The sequence of the bound pool sequence(s) is then determined, typically by cloning the DNA into a suitable sequencing vector and sequencing by standard methods. Using this

protocol, three rounds of selection resulted in approximately 80% of pool sequences with 8 randomized positions containing the binding site for ADR1 fingers. In this example only eight positions were randomized because the ninth position (G) was known to bind any residue. See, in general, Blackwell et al., Science 250:1104-1110, 1990; Wright et al., Mol. Cell. Biol. 11:4104-4110, 1991.

Within another embodiment of the invention, binding specificity of an altered DNA-binding protein can be determined by an *in vivo* assay. Briefly, a transcription promoter DNA segment containing a potential target sequence (binding site) is joined to a reporter gene so that binding of a DNA-binding protein to the target sequence will alter the transcription level of the reporter. Suitable reporter genes are those that produce a detectable phenotype in the host cell, such as genes encoding enzymes (e.g, β-galactosidase, luciferase, alcohol dehydrogenase, or amino acid biosynthetic enzymes encoded by, for example, the *LEU2* or *HIS3* genes of *S. cerevisiae*). The promoter should contain only a single potential target site, thus it is preferred to introduce the potential target sequences into a promoter that is not known to contain a binding site for a zinc finger protein. The DNA construct comprising a reporter gene operably linked to a transcription promoter segment containing a potential target sequence is introduced into a first cell (prokaryotic or eukaryotic microorganism, or cultured cell derived from a multicellar organism) according to conventional methods, the cell is cultured under conditions suitable for expression of introduced DNA, and the level of transcription is measured. It is expected that the level of transcription of the reporter gene in the first cell will be very low, unless a pre-existing, endogenous transcription factor that recognizes the binding site in the reporter construct is present in the cell. The first DNA construct and a second DNA construct

that directs expression of an altered DNA-binding protein are introduced into a second cell which is also cultured, and the level of transcription of the reporter gene is measured. Those skilled in the art will recognize that the first and second DNA constructs may be linked or unlinked, and may be autonomously replicating or may integrate into the genome of the host cell. A difference in relative transcription levels between the first and second cells is indicative of binding of the altered DNA-binding protein to the potential target sequence. A ten-fold difference in reporter transcription or activity between the first and second cells is considered a positive result. If the DNA-binding protein is a transcriptional activator, binding to the target sequence will increase expression of the reporter. Binding of a repressor DNA-binding protein will reduce expression of the reporter. This screening method can be used with chimeric DNA binding proteins, such as those having an activation domain from one protein and an altered DNA-binding domain of another protein, as well as those altered DNA binding proteins that are derived from a single parent protein. This method is also suitable for use in screening libraries of potential target sequences, in which case colonies positive for binding are isolated, and the promoter-reporter construct is sequenced to determine the sequence of the binding site. See also Sauer, R.T. (ed.), Methods Enzymol. 208, 1991.

When determining DNA binding specificity *in vitro*, it is preferred, for convenience and simplicity, to work with a polypeptide that consists essentially of the DNA-binding domain. Once binding specificity has been determined, a DNA segment encoding the DNA-binding domain can be combined with a DNA segment encoding a regulatory domain, the combined segments encoding a DNA-binding regulatory protein. When the binding specificity is to be determined by *in vivo* assay, the DNA-binding domain will,

in general, be operably linked to a regulatory domain. Most commonly, a complete DNA-binding protein of interest will be expressed in an *in vivo* assay system.

The present invention further provides DNA molecules encoding altered DNA-binding proteins. These DNA molecules can be introduced into cells according to conventional procedures such that the DNA molecules are expressed and the cells produce the altered proteins. In general, a DNA molecule encoding an altered DNA-binding protein is inserted into an expression vector, where it is operably linked to additional DNA segments that provide for its transcription. Such additional segments include promoter and terminator sequences. An expression vector may also include one or more origins of replication, one or more selectable markers, an enhancer, a polyadenylation signal, etc. Expression vectors are generally derived from plasmid or viral DNA, or may contain elements of both. The term "operably linked" indicates that the segments are arranged so that they function in concert for their intended purposes, e.g. transcription initiates in the promoter and proceeds through the coding segment to the terminator. Methods for introducing DNA into prokaryotic and eukaryotic cells and culturing the cells are well known in the art. Suitable host cells include prokaryotic cells (e.g., bacteria of the genera Escherichia and Bacillus), unicellular microorganisms (e.g., yeasts of the genera Saccharomycyes, Pichia, Schizosaccharomyces and Kluyveromyces), and cells from multicelluar organisms (e.g., mammalian, insect, avian and plant cells). See, for example, Kawasaki, U.S. Patent No. 4,599,311; Kingsman et al., U.S. Patent No. 4,615,974; Bitter, U.S. Patent No. 4,977,092; Welch et al., U.S. Patent No. 5,037,743; Murray et al., U.S. Patent No. 4,766,073; Wigler et al., Cell 14:725, 1978; Corsaro and Pearson, Somatic Cell Genetics 7:603, 1981; Graham and Van der Eb, Virology 52:456, 1973; Neumann et al., EMBO J.

1:841-845, 1982; Ausubel et al., eds., Current Protocols in Molecular Biology, John Wiley and Sons, Inc., NY, 1987; Hawley-Nelson et al., Focus 15:73-79, 1993; Hagen et al., U.S. Patent No. 4,784,950; Palmiter et al., U.S. Patent No. 4,579,821; Ringold, U.S. Patent No. 4,656,134; Foster et al., U.S. Patent No. 4,959,318; Cregg, U.S. Patent No. 4,882,279; Stroman et al., U.S. Patent No. 4,879,231; McKnight et al., U.S. Patent No. 4,935,349; Guarino et al., U.S. Patent No. 5,162,222; Bang et al., U.S. Patent No. 4,775,624; WIPO publication WO 94/06463; Sinkar et al., J. Biosci. (Bangalore) 11:47-58, 1987; Lambowitz, U.S. Patent No. 4,486,533; Sambrook et al., eds., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory, 1989; Goeddel et al., U.S. Patent No. 4,766,075; and Baird et al., U.S. Patent No. 5,155,214, which are incorporated herein by reference in their entirety.

The altered DNA-binding proteins of the present invention provide a means for regulating levels of gene transcription in eukaryotic cells. A cell that expresses an altered DNA-binding protein is used as a host for expressing one or more additional DNA molecules, particularly DNA molecules encoding proteins or polypeptides of commercial (e.g. industrial or medical) importance. Such proteins and polypeptides include growth factors, blood clotting factors, cytokines, proteases, lipases, cellulases, transglutaminases, matrix proteins, immunoglobulins, antigens, poly(amino acids), industrial enzymes, etc. A DNA molecule encoding a protein of interest is operably linked, in an expression vector, to a DNA segment encoding a transcription promoter and a binding site for the altered DNA-binding protein. Expression of the DNA-binding protein results in activation or repression of transcription of the additional DNA sequence(s). Those skilled in the art will recognize that such a system can be designed to provide

either constitutive or regulated activation or repression. Constitutive control is provided in expression systems wherein expression of the DNA binding protein is itself constitutive, thereby allowing one to increase or decrease expression of a protein of interest as necessary. For example, a low level of gene expression can be increased to an economically feasible level through constitutive co-expression of a DNA-binding transcriptional activator. In a similar manner, transcription of a protein that is detrimental to the host cell can be reduced by constitutively co-expressing a DNA-binding transcriptional repressor, thereby limiting toxicity and improving overall yield. Regulated expression of DNA-binding transcription factors using regulated promoters allows the control of expression, such as modulating expression levels to correspond with cell cycles or production cycles. For example, a cell culture can be allowed to grow to high density, at which time expression of a DNA-binding activator is initiated (or expression of a DNA-binding repressor is halted), thereby stimulating transcription of the additional DNA sequence(s) encoding one or more proteins of interest. Suitable regulated promoters in this regard include T7, *lac* and λP_L promoters for use in *E. coli*; the *GAL1*, *ADH2* and *CUP1* promoters for use in the yeast *Saccharomyces cerevisiae*; and tissue-specific promoters for use in higher eukaryotic cells.

As shown in the example which follows, *Saccharomyces cerevisiae* ADR1 was selected as a representative DNA-binding protein with which to demonstrate the methods of the present invention. ADR1 is a transcription factor of 1323 amino acids that binds to a sequence in the 5' flanking sequence of the *ADH2* gene. This binding site is known as an upstream activating sequence, or UAS. Binding of ADR1 to UAS1 in the *ADH2* gene activates *ADH2* transcription. As described below, the *ADR1* DNA sequence was altered to produce proteins

having a third zinc finger. Interactions between the three zinc fingers in the altered proteins and DNA resulted in binding of the altered proteins to sequences other than the TTGG(A/G)G recognition site of the wild-type ADR1. The methods disclosed herein are equally applicable to other DNA-binding proteins of the Cys_2-His_2 zinc finger type.

The invention is further illustrated by the following non-limiting example.

Example

Two *ADR1* polynucleotides were constructed that encode ADR1 polypeptides containing three fingers. One, designated Adr1p/F1F1F2, contains an additional finger one. The second polypeptide contains finger one motifs only and was designated Adr1p/F1F1F1 (Figure). Zinc finger genes were inserted into the expression vector pCQV (Queen, J. Mol. Appl. Genet. 2: 1-10, 1983) for expression in *E. coli* essentially as disclosed by Eisen et al. (Mol. Cell. Biol. 8: 4552-4556, 1988). An *ADR1 BamHI-SalI* fragment from pCQV ADR1 17-229, which contains the zinc finger region from amino acid 17 to amino acid 229 of ADR1, was subcloned into M13mp18 for mutagenesis. The oligonucleotide F2S1 (TGCAGAGGCC**GCATGC**ATAAGGTTTT; SEQ ID NO:10) (the antisense codon for the first Cys in the second zinc finger is in bold) was used to introduce a *SphI* site (underlined) at the end of the coding region for the first finger of ADR1. This changed the proline residue at position 133 (antisense codon GGG) to alanine (antisense codon TGC) at the amino acid level. The mutation was confirmed by DNA sequencing. The DNA fragment containing finger one and the linker sequence was cut out as a *BamHI-SphI* fragment from the mutated M13 replicative form DNA and cloned into the pCQV vector to form pCQVF1. Oligonucleotides F1S1 and F160-2 were used to copy the *ADR1* F1F2 sequence. Oligonucleotide F1S1 (AAGGTCATTT**GCATGC**GAGGTTTGTA; SEQ ID NO:11) introduced a

SphI site (underlined) containing the codon for the first Cys (bold letters) of the first finger (F1). Oligonucleotide F160-2 (GAGTCGACTTACCCTAAATTACCACTATGGATTTTTTGAGCATGTCT; SEQ. ID NO:12) introduced a *SalI* site (underlined letters) and an antisense stop codon TTA before the antisense codon for amino acid residue 160, glycine (italic letters, CCC). It also eliminated the original *SphI* site by changing C to T (bold letter). The PCR product was cut with *SphI* and *SalI*. The fragment containing F1F2 was ligated into *SphI*- and *SalI*-digested pCQVF1 to form pCQVAdr1F1F1F2. The F1 sequence was also cloned using oligonucleotides F1S1 and F2S1 as primers for PCR. The PCR product was digested with *SphI*, and two copies were cloned in tandem into *SphI*-digested pCQVF1 to form pCQV/ADR1F1F1F1.

Zinc finger proteins were expressed in *E. coli* MC1061 using pCQV as the expression vector. Protein extracts were prepared essentially as described by Eisen et al. (ibid) and boiled for 15 minutes in 200 mM KCl, 100 μm $ZnCl_2$, 5mM DTT, 10% glycerol, 25 mM HEPES pH 7.5, and 5 mM $MgCl_2$ ($A_{200}Z$ buffer). The expression of the zinc finger proteins was confirmed by Western blotting using anti-ADR1 finger serum as described by Eisen et al. (ibid.) Zinc finger proteins accounted for approximately 60% of the total protein after the heat treatment as determined by SDS gel electrophoresis and titration of the DNA binding activity using DNA probes.

Adr1p/F1F1F2 was tested with several different oligonucleotide probes that contained two copies of the predicted binding site, TTG G(A/G)G G(A/G)G, in inverted orientation, as found in the wild-type *ADR1* UAS1 (Thukral et al. Mol. Cell. Biol. 11:1566-1577, 1991).

Gel retardation assays were done essentially as described by Thukral et al. (ibid.). Probes were derived from the oligonucleotides containing UAS1 and UAS2 sequences with changes at the binding sites and labeled

with γ^{32}P-ATP. A DNA probe with the DNA binding site of TTG GGG GGG was synthesized from the template oligonucleotide F1²F2 (GTTCTCCCCCCCAACTTATAAGTTGGGGGGGGATGCCCCGGTGTTCCGGCAGAGGAGA GGTAC; SEQ ID NO:13) and a primer (SK-primer CTCTCCTCTGCCGGAACA; SEQ ID NO:14). Other binding sites were derived by modification of the binding sites (inverted repeat, underlined). In the DMS interference experiment, oligonucleotide GTCATGACTCAGGTAAGTTGGGGGGGGATGCCCCGGTGTTCCGGCAGAGGAGAGGTAC (SEQ ID NO:15) (zinc finger protein binding site underlined) was labeled with γ^{32}P-ATP using polynucleotide kinase and made double stranded using SK-primer as the primer with DNA polymerase. The probe was purified from a 12% polyacrylamide gel. The free and protein-bound probes were isolated after separating them using the DNA migration assay.

For the DNA migration assay, 12 nM of zinc finger protein was mixed with 1-32 nM of DNA probe in a 20 µl reaction volume. The DNA added to each reaction was quantitated using absorbance at 260 nm. The reaction mixture was placed on ice for 15 minutes, then held at room temperature for 5 minutes. The mixture was then electrophoresed on a 5% acrylamide gel under non-denaturing conditions at 200 volts. The gel was autoradiographed and scanned on a densitometer or using a program NIH image. K_{app} was calculated as described in the equation: $[CD]/[D] = -K_{app}[CD] + K_{app}[C0]$ (Baker et al., J. Biol. Chem. 261: 5275-5282, 1986). At least three independent assays were done, and the standard deviations were less than 25%. Results are shown in Table 2.

Table 2: Binding of F1F1F2 finger to different DNA sites

DNA site	K_{app} (M^{-1})
TTG GGG GGG	2×10^8
TTG GAG GAG	8×10^7
TTc GGG GGG	3×10^7
TTG GGc GGG	1×10^7
TTG GGG GGc	1×10^6
TTc GGG GGc	2×10^7
TTc GGc GGG	-
TTG GGc GGc	-
TTG GAG	-
TTG GAG	6×10^7*

Affinity of the three-finger F1F1F2 binding to various DNA sites was determined by similar DNA gel retardation assay. Kapp was given: $[CD]/[D] = -K_{app}[CD] + K_{app}[C_0]$ [CD]: the concentration of DNA-protein complex; [D]: the concentration of free DNA; $[C_0]$: the concentration of total zinc finger protein. * denotes DNA binding affinity of Adr1p/F1F1 protein. Concentrations were determined by quantitative desitometry of autoradiographs.

The three-finger protein bound well to TTG G(A/G)G G(A/G)G oligonucleotides, forming complexes containing both one (CI) and two molecules (CII) of Adr1p/F1F1F2 as did the two-finger protein on its binding site. Quantitation of DNA binding showed that it did not

bind at a detectable level to the binding site of the wild-type two finger protein, TTG GAG.

To assess the contributions of the individual fingers, single and double mutations were introduced at positions in the binding site known to be important for binding to the wild-type protein. The results of quantitative DNA binding suggest that each finger contributes to the binding but to different degrees (Table 2). The contribution by the N-terminal finger seemed least important and the contribution by the middle finger seemed most important, even though these fingers are identical.

Contacts between Adr1p/F1F1F2 and its optimal binding site were examined by methylation interference. An oligonucleotide (GTCATGACTCAGGTAAGTTGGGGGGGATGCCCGGTG-TTCCGGCAGAGGAGAGGTAC; SEQ ID NO:15) (binding site underlined) was labeled with [γ-P^{32}]-ATP using polynucleotide kinase and made double stranded using SK-primer (CTCTCCTCTGCCGGAACA; SEQ ID NO:14) and DNA polymerase. The probe was purified from a 12% polyacrylamide gel and treated with 0.5 μl dimethyl sulfate at room temperature for 2 minutes. The free and protein-bound probes were isolated after separating them using a DNA migration assay as described above. The DNA sequence was determined as described by Maxam and Gilbert, Methods Enzymol. 65: 499-560, 1980. Modification of the guanines in the sequence TTG GGG GGG interfered with binding as shown in Table 2, confirming the importance of Arg and His contacts with guanine, and suggesting that the three-finger protein contacts these residues through the major groove as expected.

The specificity of Adr1/F1F1F2 was compared to that of the wild-type (Adr1/F1F2). To compare the specificities of the two proteins, DNA binding assays were carried out in the presence of increasing amounts of "random" DNA. The non-specific association constant of Adr1/F1F2 was found to be appreciably higher than that of

Adr1/F1F1F2. The average value of the Specificity, defined as $K_{specific}/K_{non-specific}$, was about 2 x 10^2 for Adr1/F1F2 and about 2 x 10^3 for Adr1/F1F1F2 (Table 3). Thus, Adr1/F1F1F2 has about a ten-fold greater specificity that the wild-type sequence.

Table 3: Specificity of binding of Adr1/F1F2 and Adr1/F1F1F2

Protein	Specific Binding[1]	Non-specific Binding[1]	Specificity Ratio
Adr1/F1F2	2.0 x 10^8	1 x 10^6	200
Adr1/F1F1F2	4.5 x 10^8	2 x 10^5	2000

[1]Binding constant, K, in units of M^{-1}

ADR1 containing three finger one motifs did not efficiently recognize its predicted binding site, G(A/G)G G(A/G)G G(A/G)G. The binding site for this protein was determined by a binding site selection and amplification assay (data not shown). The binding sites that were selectively amplified contained the consensus sequence NG(G/T/A) G(T/C)G GGG (where N represents any nucleotide). This result was reconfirmed by gel mobility assay using different DNA sequences as probes. In binding this sequence, Adr1p(F1F1F1) would be able to make 7 of the 9 predicted contacts between Arg or His in the fingers and guanine in the DNA. Since the consensus sequence does not contain a repetitive triplet motif, the three identical fingers must be contacting different subsites. Although it is not possible to unambiguously determine from these data which finger is contacting which base pairs in this binding site, it seems likely that the N-terminal finger is contacting a predicted subsite, GGG, that the middle finger is contacting GCG, and that the C-terminal finger is contacting NG(G/T/A). If this interpretation is correct, the middle subsite contains a C at the central position in place of G or A, a change that prevents binding by wild-type *ADR1*, and the subsite for the C-

terminal finger contains only two positions showing base specificity.

To assess the ability of Adr1p/F1F1F2 to function *in vivo* the DNA binding domain was fused to the Herpes simplex virus VP16 activation domain (Regier et al., Proc. Natl. Acad. Sci. USA 90: 883-887, 1993; Sadowski et al., Nature 335: 563-564, 1988). An analogous construct was made containing the normal two-finger domain. The fusion proteins were expressed in *Saccharomyces cerevisiae* containing a plasmid bearing a reporter gene, and β-galactosidase assays were performed to monitor their activity. Yeast cells were grown in synthetic medium with ethanol as the carbon source for 12 hours and harvested for the assay of β-galactosidase as described by Cheng et al., Mol. Cell. Biol. 14: 3842-3852, 1994. pRSADR1-VP16 is a centromere plasmid based on pRS314 (Guarente et al., Proc. Natl. Acad. Sci USA 79: 7410-7414, 1982) containing the *TRP1* gene as the marker and the *ADR1* gene fused to VP16 at amino acid codon 172. The VP16 gene codes for its C-terminal activation domain from amino acid 413 to amino acid 490. The *EcoRI-BclI* fragment from pCQVF1F1F2, which contains the three-finger fragment, was cloned into pRSADR1-VP16 to form pRSF1F1F2-VP16. Mutations in the middle F1 and F2 were introduced by PCR. The mutant DNAs were cloned in M13mp18 as described by Thukral et al., Mol. Cell. Biol. 12: 2704-2792, 1992 and Thukral et al., Proc. Natl. Acad. Sci USA 88: 9188-9192, 1991 and used as templates. Oligonucleotides F1S1 and F160-4 were used for PCR. Oligonucleotide F160-4 (GGGAAACTGCAGCCCCTAAATTACCACTATGGATTTTTTGAGCATGTCT; SEQ. ID NO:16) introduced a *PstI* site (underlined) at amino acid codon 160 (in bold letters), which is in frame with sequence in pRSF1F1F2-VP16. F160-4 also eliminated the original *SphI* site (in bold and underlined letters) around the amino acid 149 (arginine) codon without changing the amino acid sequence. PCR products were digested with *SphI*

and *Pst*I and cloned into pRSF1F1F2-VP16 to form mutated zinc fingers.

The three-finger protein, Adr1p/F1F1F2-VP16, was an efficient transcriptional activator, and its function *in vivo* reflected its binding activity determined *in vitro* (Table 4). Its activity differed both qualitatively and quantitatively from that of the two-finger protein. Surprisingly, the three-finger protein activated a reporter containing a single binding site, unlike the two finger protein, which required an inverted repeat at its binding site in order to activate transcription. It was also unexpected that activation promoted by the three-finger fusion protein was about 50 times higher than that promoted by the two-finger fusion protein.

Table 4: Gene Activation by three zinc finger activator

	β-Gal Activity (Miller Units)					
	Adr1p Activator					
Sites in reporter*	F1F2 -VP16	F1F1F2 -VP16	F1F1 (R115A)-F2-VP16	F1F1 (H118T)-F2-VP16	F1F1F2 (R143A) -VP16	–
Inverted repeat						
TTG GGG GGG	6.6	163.0	5.8	14	99	6.2
TTG GAG GAG	NT	23.2	1.0	4.3	0.8	0.7
TTG GAG	9.6	0.2	0.3	1.3	0.3	0.3
TTc GGc GGG	NT	0.5	NT	0.8	0.7	0.7
Single site						
TTG GGG GGG	5.9	100.0	14	24	43	16
TTG GCG GGG	NT	3.6	7.3	203	4.9	6.8
TTG GAG	1.7	0.9	NT	NT	NT	1.3
–	NT	NT	0.3	0.2	NT	NT

Table 4 Continued

*Construction of reporter genes: pLGK is a 2μ plasmid with a *URA*3 gene as the marker and a *lacZ* gene fused to the *CYC*1 promoter sequence (Guarente et al., ibid.). The oligonucleotides that contain a single or an inverted repeat binding site for the zinc finger were double stranded and cloned into the *KpnI-XhoI* site in pLGK. The sequences for the single or inverted repeat DNA binding sites were the same as used in DNA migration assays. Deletion of the *KpnI-XhoI* fragment in pLGK removes all of the UAS element in the *CYC*1 promoter.

Activation through a single site could be attributable to the VP16 activation domain on the three-finger fusion protein even though the two-finger fusion protein did not show this property. To test this possibility the three-finger DNA binding domain was put into the context of full length Adr1p (containing 1323 amino acids) by substituting the zinc finger domain. The resulting Adr1p(1323)/F1F1F2 activated the reporter gene with its binding site in the UAS element present either as an inverted repeat or as a single site (data not shown). Thus, the ability to activate transcription through a single site is conferred by the DNA binding domain, not by the activation domain.

Mutations were introduced into the middle (F1) or the C-terminal finger (F2) in the yeast expression vector containing *ADR*1/F1F1F2-VP16. A mutation of His 118 to Thr in the middle finger conferred a new DNA binding specificity on the protein, allowing it to activate a reporter gene with the sequence TTG GCG GGG in its promoter (F1F1(H118T)F2; Table 4). When present in wild-type Adr1p, the analogous mutation allowed the protein to bind to and activate from TTG GCG (Thukral et al., 1992, ibid). This result further supports the idea that Adr1p/F1F1F2 binds DNA in the predicted manner, with F1 contacting a triplet, G(A/G)G, in the three-finger protein as it does in the two-finger protein. Moreover, Adr1p/F1F1(H118T)F2-VP16 activated through a single site, as did the wild-type version of this three-finger protein, suggesting that it binds tightly *in vivo*.

Mutants representing loss-of-function alleles had phenotypes that depended on their locations. The R143A mutation in finger 2 reduced activity less than two-fold, while the R115A mutation in the middle finger reduced activity about thirty-fold (F1F1F2(R143A)-VP16; F1F1(R115A)F2-VP16; Table 4). Adr1p with just two fingers was completely defective with either of these mutations (Thukral et al., 1991, ibid). These data support the DNA binding data that suggested that the middle finger is more important than the N- or C-terminal fingers.

These results show that the position of a finger can influence both its specificity and is contribution to binding affinity. The same finger contributed differently to the overall affinity depending on whether it was the N-terminal or middle finger in Adr1p/F1F1F2. In Adr1p/F1F1F1 the binding sites for each finger appeared to be different, even though the fingers were identical. Despite these context effects, addition of a third zinc finger to ADR1 in each case altered binding specificity, and each of the three fingers made sequence-specific contacts with the DNA.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: YOUNG, ELTON T.
CHENG, CHENG

(ii) TITLE OF INVENTION: METHODS FOR PREPARING DNA-BINDING PROTEINS

(iii) NUMBER OF SEQUENCES: 16

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: ZYMOGENETICS, INC.
(B) STREET: 1201 EASTLAKE AVENUE EAST
(C) CITY: SEATTLE
(D) STATE: WA
(E) COUNTRY: USA
(F) ZIP: 98102

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: PARKER, GARY E.
(B) REGISTRATION NUMBER: 31,648
(C) REFERENCE/DOCKET NUMBER: 92-01

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 206-442-6600
(B) TELEFAX: 206-442-6678

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 2
(D) OTHER INFORMATION: /label= Xaa
/note= "This amino acid can be Tyr or Phe"

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 3
(D) OTHER INFORMATION: /note= "This amino acid can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 5..8
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 10..12
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 14..18
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 20..21

(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 23..26
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Pro Xaa Xaa Cys Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Phe Xaa Xaa Xaa
1 5 10 15

Xaa Xaa Leu Xaa Xaa His Xaa Xaa Xaa Xaa His Thr Gly Glu Lys
20 25 30

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 2
(D) OTHER INFORMATION: /note= "This amino acid can be Tyr or Phe."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 3
(D) OTHER INFORMATION: /note= "This amino acid can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 5..8

(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 10..12
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 14..18
(D) OTHER INFORMATION: /note= "These nino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 20..21
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 23..25
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Pro Xaa Xaa Cys Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Phe Xaa Xaa Xaa
1 5 10 15

Xaa Xaa Leu Xaa Xaa His Xaa Xaa Xaa His Thr Gly Glu Lys
20 25 30

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 2
(D) OTHER INFORMATION: /note= "This amino acid can be Tyr or Phe."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 3
(D) OTHER INFORMATION: /note= "This amino acid can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 5..7
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 9..11
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 13..17
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 19..20
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 22..24
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Pro Xaa Xaa Cys Xaa Xaa Xaa Cys Xaa Xaa Xaa Phe Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Leu Xaa Xaa His Xaa Xaa Xaa His Thr Gly Glu Lys
20 25

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 2
(D) OTHER INFORMATION: /note= "This amino acid can be Tyr or Phe."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 3
(D) OTHER INFORMATION: /note= "This amino acid can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 5..6
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 8..10
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 12..16
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 18..19
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 21..23
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Pro Xaa Xaa Cys Xaa Xaa Cys Xaa Xaa Xaa Phe Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Leu Xaa Xaa His Xaa Xaa Xaa His Thr Gly Glu Lys
20 25

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 2
(D) OTHER INFORMATION: /note= "This amino acid can be Tyr or Phe."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 3
(D) OTHER INFORMATION: /note= "This amino acid can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 5..7
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 9..11
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 13..17
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 19..20
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 22..25
(D) OTHER INFOPMATION: /note= "These amino acids can be any amino acid."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Pro Xaa Xaa Cys Xaa Xaa Xaa Cys Xaa Xaa Xaa Phe Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Leu Xaa Xaa His Xaa Xaa Xaa Xaa His Thr Gly Glu Lys
20 25 30

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 2
(D) OTHER INFORMATION: /note= "This amino acid can be Tyr or Phe."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 3
(D) OTHER INFORMATION: /note= "This amino acid can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 5..6
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 8..10
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 12..16
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 18..19

(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 21..24
(D) OTHER INFORMATION: /note= "These amino acids can be any amino acid."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Pro Xaa Xaa Cys Xaa Xaa Cys Xaa Xaa Xaa Phe Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Leu Xaa Xaa His Xaa Xaa Xaa Xaa His Thr Gly Glu Lys
20 25

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGNNGGNAGG ANNGGGNGGN NNANNNG 27

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GAYAAGATAA GATAA 15

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 2
(D) OTHER INFORMATION: /note= "This amino acid can be Gly or Asn."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Thr Xaa Glu Lys Pro
1 5

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TGCAGAGGCC GCATGCATAA GGTTTT 26

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AAGGTCATTT GCATGCGAGG TTTGTA 26

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAGTCGACTT ACCCTAAATT ACCACTATGG ATTTTTTGAG CATGTCT 47

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 61 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTTCTCCCCC CCAACTTATA AGTTGGGGGG GATGCCCGGT GTTCCGGCAG AGGAGAGGTA 60

C 61

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTCTCCTCTG CCGGAACA 18

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTCATGACTC AGGTAAGTTG GGGGGGATGC CCGGTGTTCC GGCAGAGGAG AGGTAC 56

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GGGAAACTGC AGCCCCTAAA TTACCACTAT GGATTTTTTG AGCATGTCT 49

Claims

We claim:

1. A method for preparing a DNA-binding protein having altered binding specificity comprising:

selecting a parent DNA-binding protein comprising first and second Cys_2-His_2 zinc fingers, wherein the DNA binding specificity of said parent protein is known;

adding an additional Cys_2-His_2 zinc finger to said parent protein to produce an altered DNA-binding protein; and

determining the DNA binding specificity of said altered DNA-binding protein, wherein the binding specificity of the altered protein is a result of interactions between nucleotides in a target sequence and amino acid residues in each of said first, second, and additional zinc fingers.

2. A method according to claim 1 wherein the determining step comprises measuring electrophoretic mobility of a complex of said altered DNA-binding protein and a DNA molecule.

3. A method according to claim 1 wherein the determining step comprises:

preparing a mixture comprising the altered DNA-binding protein and a DNA molecule comprising a predicted binding site under conditions suitable for formation of protein-DNA complexes; and

measuring complex formation between the altered DNA-binding protein and the DNA molecule.

4. A method according to claim 3 wherein the measuring step comprises measurement of electrophoretic mobility of a complex of the altered DNA-binding protein and the polynuclectide molecule.

5. A method according to claim 1 wherein the determining step comprises:

preparing a mixture comprising the altered DNA-binding protein and a plurality of DNA molecules under conditions suitable for complex formation between the altered protein and a target DNA molecule;

isolating a complex of the altered DNA-binding protein and a target DNA molecule;

amplifying the target DNA molecule; and

determining the sequence of a binding site for said altered DNA-binding protein in said target DNA molecule.

6. A method according to claim 1 wherein the determining step comprises:

culturing a first cell into which has been introduced a first DNA construct comprising a reporter gene operably linked to a transcription promoter segment containing a potential target sequence;

culturing a second cell into which has been introduced said first DNA construct and a second DNA construct that directs expression of said altered DNA-binding protein;

measuring transcription of said reporter gene in said first and second cells, wherein a difference in relative transcription levels is indicative of binding of said altered DNA-binding protein to said potential target sequence.

7. A method according to claim 1 wherein said additional zinc finger is a duplicate of one of said first and second zinc fingers.

8. A method according to claim 1 wherein said parent DNA-binding protein is a wild-type *Saccharomyces cerevisiae* ADR1 protein or MIG1 protein.

9. A method according to claim 1 wherein said parent DNA-binding protein is *Saccharomyces cerevisiae* ADR1 having a mutation in one of said first or second fingers that changes DNA binding specificity as compared to wild-type *Saccharomyces cerevisiae* ADR1.

10. A method according to claim 1 wherein said target sequence is from 9 to 15 nucleotides in length.

11. A method according to claim 1 wherein said altered DNA-binding protein has three, four or five Cys_2-His_2 zinc fingers, each of which interacts with said target sequence.

12. A method according to claim 1 wherein said parent DNA-binding protein has two, three or four Cys_2-His_2 zinc fingers.

13. A DNA-binding protein comprising first and second Cys_2-His_2 zinc fingers which has been modified to contain an additional Cys_2-His_2 zinc finger, wherein said DNA-binding protein binds to a binding site in DNA, wherein said binding is a result of interactions between nucleotides in said binding site and amino acid residues in each of said first, second, and additional zinc fingers.

14. A DNA-binding protein according to claim 13 which contains only three zinc fingers.

15. A DNA-binding protein according to claim 13 wherein said additional zinc finger is a duplicate of one of said first and second zinc fingers.

16. A DNA-binding protein according to claim 13 which is a *Saccharomyces cerevisiae* protein selected from the group consisting of ADR1 and MIG1.

17. A DNA-binding protein according to claim 13, wherein said protein is a *Saccharomyces cerevisiae* ADR1 protein which has been modified to contain a third Cys_2-His_2 zinc finger, wherein said ADR1 protein binds to a binding site other than TTGG(A/G)G.

18. An ADR1 protein according to claim 17 wherein said third zinc finger is a duplicate of a zinc finger in a wild-type ADR1 protein.

19. An ADR1 protein according to claim 17 wherein said third zinc finger is that of a DNA-binding protein other than ADR1.

20. An ADR1 protein according to claim 17 which is further modified to contain a fourth Cys_2-His_2 zinc finger, wherein said fourth zinc finger alters binding specificity of the ADR1 protein.

21. A cultured eukaryotic cell into which has been introduced a gene encoding a DNA-binding protein comprising first and second Cys_2-His_2 zinc fingers, wherein said gene has been modified so that said DNA-binding protein contains an additional Cys_2-His_2 zinc finger, wherein said DNA-binding protein binds to a binding site in DNA, and wherein said binding is a result of interactions between nucleotides in said binding site and amino acid residues in each of said first, second, and additional zinc fingers.

22. A cultured eukaryotic cell according to claim 21 which is a fungal cell.

23. A cultured eukaryotic cell according to claim 22 which is a yeast cell or an Aspergillus cell.

24. A cultured eukaryotic cell according to claim 21 wherein said DNA-binding protein is a modified *S. cerevisiae* ADR1 or MIG1 protein.

25. A cultured eukaryotic cell according to claim 21 wherein said DNA-binding protein contains only three zinc fingers.

26. A cultured eukaryotic cell according to claim 21 wherein said additional zinc finger is a duplicate of one of said first and second zinc fingers.

27. A cultured eukaryotic cell according to claim 21 into which has been introduced a first DNA segment encoding a polypeptide of interest operably linked to a second DNA segment comprising a transcription promoter and a binding site for said DNA-binding protein, wherein binding of said DNA-binding protein to said binding site stimulates transcription of said first DNA segment.

28. A cutured eukaryotic cell according to claim 21 wherein said cell is a yeast cell, said gene is an *ADR1* gene, and said binding site is a site other than TTGG(A/G)G.

29. A yeast cell according to claim 28 which is a *Saccharomyces cerevisiae* cell.

30. A yeast cell according to claim 28 into which has been introduced a first DNA segment encoding a polypeptide of interest operably linked to a second DNA segment comprising a transcription promoter and a binding site for said DNA-binding protein, wherein binding of said DNA-binding protein to said binding site stimulates transcription of said first DNA segment.

31. A method for preparing a polypeptide of interest comprising:

(a) culturing a yeast cell into which has been introduced:

an *ADR1* gene modified to encode a protein containing a third Cys_2-His_2 zinc finger, wherein said *ADR1*-encoded protein binds to a binding site other than TTGG(A/G)G as a result of interactions between nucleotides in the binding site and amino acid residues in said third zinc finger; and

a first DNA segment encoding a polypeptide of interest operably linked to a second DNA segment comprising a transcription promoter and a binding site for said *ADR1*-encoded protein, wherein binding of said *ADR1*-encoded protein to said binding site stimulates transcription of said first DNA segment,

under conditions suitable for expression of said *ADR1* gene and said first DNA segment; and

(b) isolating the polypeptide of interest from said yeast cell.

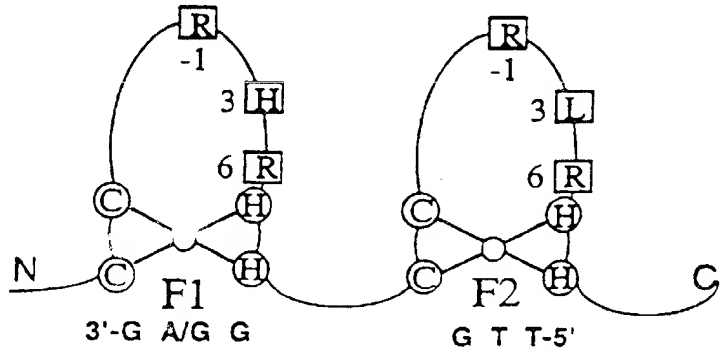
32. A cultured eukaryotic cell into which has been introduced a a gene encoding a chimeric transcription factor, wherein said transcription factor comprises a *S. cerevisiae* ADR1 DNA-binding domain modified to contain a third Cys_2-His_2 zinc finger, wherein said ADR1 DNA-binding domain binds to a binding site other than TTGG(A/G)G as a result of interactions between nucleotides in the binding site and amino acid residues in said third zinc finger, and wherein said chimeric transcription factor further comprises a non-ADR1 transcription activation or repression domain operably linked to said DNA-binding domain.

33. A cultured eukaryotic cell according to claim 32 wherein said non-ADR1 domain is a transcription activation domain from *S. cerevisiae* GAL4, *S. cerevisiae* GCN4, human SP1, mouse SP1, or Herpes simplex virus VP16.

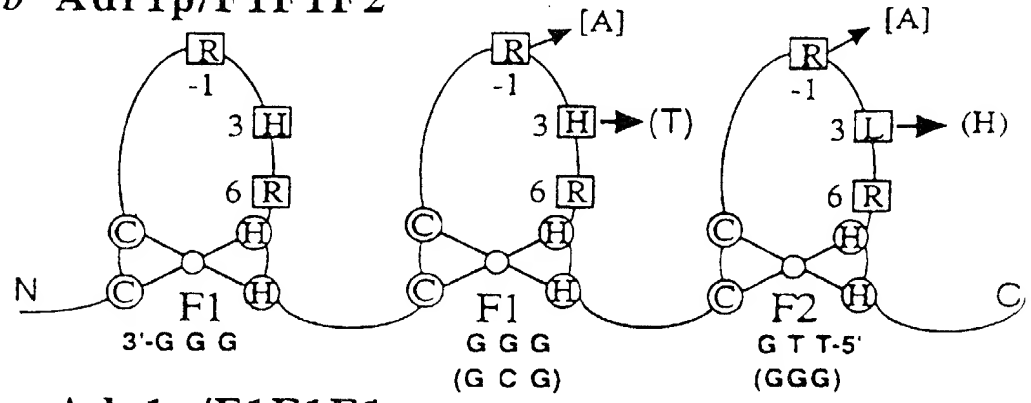
34. A cultured eukaryotic cell according to claim 32 wherein said non-ADR1 domain is a steroid receptor family transcription activation domain.

35. A cultured eukaryotic cell according to claim 32 which is a yeast cell.

a **Adr1p/F1F2**



b **Adr1p/F1F1F2**



c **Adr1p/F1F1F1**